

HENP Grand Challenge Project*

D. Olson†, A. Vaniachine†, J. Yang‡

The High Energy and Nuclear Physics Data Access Grand Challenge project has developed an optimizing storage access software system that was prototyped at RHIC. It is currently undergoing integration with the STAR experiment in preparation for data taking that starts in mid-2000. The system was exercised during the RHIC Mock Data Challenges and tested under conditions designed to characterize scalability, up to 100 simultaneous queries, 10 M events across 7 event components. The system coordinates the staging of "bundles" of files from the HPSS tape system, so that all the needed components of each event are in disk cache when accessed by the application software. The initial implementation interfaced to the Objectivity/DB. In this latest version, it evolved to work with arbitrary files and use CORBA interfaces to the tag database and file catalog services.

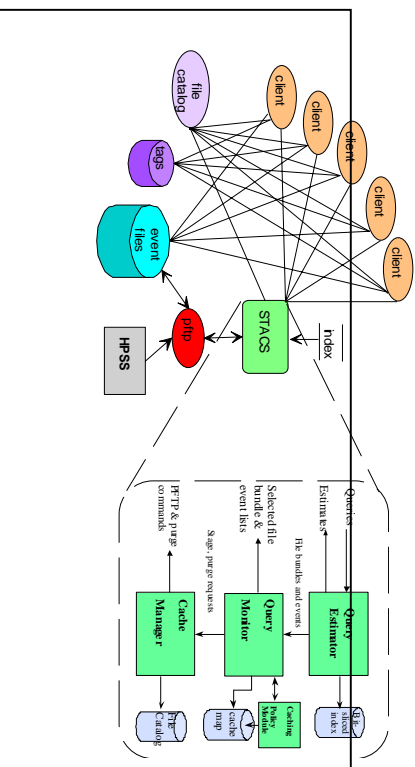


Figure 1. Illustration of system software architecture with exploded view of Storage Access Coordination System (STACS)

STACS has 3 main components: 1) The Query Estimator (QE), that uses the index to determine what files and what events are needed to satisfy a given range query. 2) The Query Monitor (QM), that keeps track of what queries are executing at any time, what files are cached on behalf of each query, what files are not in use but are still in cache, and what files still need to be

cached. The Query Monitor consults an additional module, called the Caching Policy module, 3) The Cache Manager, that is responsible for interfacing to the mass storage system (HPSS) to perform all the actions of staging files to and purging files from the disk cache.

In the multi-component event model, each event is partitioned into several components, such as "tracks", "hits", and "raw". We introduced the term "file bundle" to refer to the ordered set of files, one for each component, that needs to be in cache at the same time to process events whose components are in these files.

Scalability testing was done for the purpose of finding areas where the system can potentially break as the number of events, files, and queries increases. A test dataset was set up for about 10 million events, each partitioned into 7 components, organized into some 4700 files, totaling about 1.6 TB. All tests were successful, and the system has been running for up to a week without any failure in our test.

The STAR experiment has adopted the MySQL database to keep records of data files and their production history. Records for all files are kept in one database table – fileCatalog. During the event reconstruction a set of structures containing selected event information is saved as the tag component of the event (about 500 in STAR). The STAR event tags consist of overall event summary tags, daq/online tags, and a set of useful physics tags. These tags are used to construct the index used by STACS.

Footnotes and References

- *<http://www-mc.lbl.gov/GC/>
- † RNC Program, LBNL/NSD
- ‡ UCLA & LBNL